Benefits of using Apache Storm on HDInsight

Benefits of using Apache Storm on HDInsight.
In this video,
we'll explain why you should use Apache Storm on HDInsight.
Then we'll highlight the common use cases for
real-time analytics.
Apache Storm on HDInsight is a managed cluster integrated into
the Azure environment.
It provides the following key benefits.
It performs as a managed service with an SLA of 99.9% up time,
use the language of your choice.
It provides support for Storm components written in Java, C#,
and Python.
It also supports a mix of programming languages,
read data using Java, then process it using C#.
Please note that C# topologies are only supported
on Windows based HDInsight clusters.
Analyze millions of messages per second.
The HDInsight team runs a lab test that processes 1 million
messages per second ingested from Event Hub.
Use the Trident Java interface to create Storm topologies
that support exactly wants processing in messages,
transactional data store persistence and
a set of common Stream Analytics operations.
Storm on HDInside includes built-in scale up and
scale down features.
Scale in HDInside cluster with no impact
to currently running Storm Topologies.
Integrate with other Azure services, including Event Hub,
Azure Virtual Network, SQL Database, Blob Storage and
DocumentDB.
Then combine the abilities of multiple
HDInsight clusters by using Azure Virtual Network.
Create analytic pipelines that use HDInsight, HBase or
Hadoop clusters.
You can develop and deploy Storm applications completely
from Visual Studio using the following steps.
Select the Storm application project template.
Create a basic topology with a simple spout and bolt.
Publish the new project to Azure from the Visual Studio.
You right-click on the project and the Solution Explorer and
select submit to Storm on HDInsight.
And verify it's all working correctly from the Storm
topologies video, then submit the topology.
Monitor the topology with Visual Studio and finally test and
troubleshoot the topology and view error information.
The following are some common scenarios for
which you might use Apache Storm on HDInsight.
Internet of Things or IoT, it may seem obvious or inevitable

but many companies are embracing the Internet of Things for the following reasons.
McKinsey Global Institute reports that IoT business will reach $6.2 trillion in revenue by 2025.
More and more objects are becoming embedded with sensors that communicate real-time data to data centers and networks for processing.
It's also being used for real-time fraud detection by many companies.
Another use case is for real-time social analytics.
Other use cases are Extract, Transform and Load (ETL), Network monitoring, search, and mobile engagement.
Some of the companies that are using Storm are Groupon.
At Groupon they use Storm to build real-time data integration systems.
Storm helps them analyze, clean, normalize, and resolve large amounts of non-unique data points with low latency and high throughput.
The Weather Channel, at The Weather Channel, they use several Storm topologies to ingest and persist weather data.
Each topology is responsible for fetching one dataset from an internal or external network, i.e., the Internet, reshaping the records for use by their company and persisting the records to relational databases.
It is particularly useful to have an automatic mechanism for repeating attempts to download and manipulate the data when there is a hiccup.
Twitter, Storm powers a wide variety of Twitter systems, ranging in applications from discovery, real-time analytics, personalization, search, revenue optimization, and many more.
Storm integrates with rest of Twitter's infrastructure, including database systems Cassandra, Memcached, etc., the messaging infrastructure,
Mesos and the monitoring and alerting systems.
Storm's isolation scheduler, makes it easy to use the same cluster both for production applications and in development applications.
And it provides a sane way to do capacity planning.
Yahoo, Yahoo is developing a new generation platform that enables the convergence of big data and low latency processing.
While Hadoop is their primary technology for batch processing, Storm empowers stream and micro-batch processing of user events, content feeds, and application logs.
Spotify, Spotify serves streaming music to over 10 million subscribers and 40 million active users.
Storm powers a wide range of real-time features at Spotify, including music recommendation, monitoring, analytics and adds targeting.
Together with Kafka, Memcash, Cassandra and NettyZMPT based messaging, Storm enables them to build

low latency, fault tolerant, distributed systems with ease.
Alibaba, Alibaba is the leading B2B e-commerce
website in the world.
They use Storm to process the application log and the data
change in the database to supply real-time stats for data apps.
Klout, Klout helps everyone discover and be recognized for
their influence by analyzing engagement with their
content across social networks.
Their analysis powers a daily Klout score
on a scale of 1 to 100 that shows
how much influence social media users have, and on what topics.
They are using Storm to develop a real-time scoring and
moments generation pipeline.
Leveraging Storm's intuitive trident abstraction,
they are able to create complex topologies which stream data
from our network collectors, the Acafa, and process and
written out to HDFS.